JOURNAL OF MULTIVARIATE ANALYSIS **60**, 123–147 (1997)
ARTICLE NO. MV961647

On Consistency in Nonparametric Estimation under Mixing Conditions

A. Irle*, †

University of Kiel, Kiel, Germany

In this paper a method for obtaining a.s. consistency in nonparametric estimation is presented which only requires the handling of covariances. This method is applied to kernel density estimation and kernel and nearest neighbour regression estimation. It leads to conditions for a.s. consistency which relax known conditions and include long-range dependence. © 1997 Academic Press

View metadata, citation and similar papers at core.ac.uk

1. INTRODUCTION

The question of consistency in nonparametric estimation under mixing conditions has found considerable recent interest. We point out the papers by Collomb (1984), Roussas (1988) and the lecture notes by Györfi, Härdle, Sarda and Vieu (1989) for results on consistency for general mixing processes. In these references exponential probabilistic inequalities are used to show consistency in the sense of complete convergence from which of course via the Borel-Cantelli-Lemma consistency in the usual sense of a.s. convergence follows. It may be guessed that a.s. consistency alone can be proved under substantially less restrictive conditions than complete convergence.

We shall show in this paper that this guess is correct and present a method by which a.s. consistency may be obtained directly. This method was used by Etemadi (1981) to prove the strong law of large numbers for i.i.d. random varibles without the use of an inequality of Kolmogorov type. Distinguishing between short-range dependence as defined by summability of the mixing coefficients and long-range dependence this device works

Received April 10, 1995; revised August 15, 1996.

AMS 1991 subject classifications: 62G07, 62G20.

Key words: Nonparametric estimation, consistency, mixing, kernel estimation, nearest neighbour estimation.

* Work done as Visiting Fellow to the Centre for Mathematics and Its Applications of the Australian National University and supported by a grant of the Volkswagenstiftung.

† Postal address: Mathematisches Seminar, University of Kiel, Ludewig-Meyn-Strasse 4, D-24098 Kiel, Germany.

0047-259X/97 $25.00

Copyright © 1997 by Academic Press
All rights of reproduction in any form reserved.

equally well for long-range dependence. On the probabilistic side it only invokes the handling of covariances so that it is easily adapted to different types of mixing conditions. Here we shall consider ϕ-mixing and α-mixing coefficients. As a minor byproduct we relax the assumption of stationarity.

The main tool is a general convergence result given in Section 2 which obtains the convergence

$$1/(na(n)) \sum_{i=1}^{n} (K_n(Z_i) - EK_n(Z_i)) \to 0 \quad \text{a.s.}$$

under suitable assumptions for positive constants $a(n)$, real-valued mappings K_n and mixing random variables Z_n, $n = 1, 2, \ldots$. As an example of the conditions used the essential condition in the case of ϕ-mixing states

$$\sum_{i=1}^{\infty} \sum_{j=1}^{i} \phi(j)/(i^2 a(i)) < \infty,$$

which under short-range dependence becomes

$$\sum_{i=1}^{\infty} 1/(i^2 a(i)) < \infty$$

so no longer containing any interaction between constants and mixing coefficients in this case, whereas for long-range dependence such interaction occurs.

In Section 3 we consider kernel density estimation and kernel regression estimation in d-dimensional space. In the ϕ-mixing case the essential condition is stated as

$$\sum_{i=1}^{\infty} \sum_{j=1}^{i} \phi(j)/(i^2 h(i)^d) < \infty$$

with $h(n)$ denoting the bandwidth of the kernel estimator.

Nearest neighbour regression estimation is treated in Section 4, the number of nearest used being denoted by $k(n)$. For points x in the domain of the regression function which are continuity points of the underlying distribution, an approach initiated by Moore and Yackel (1977) and using suitable related kernel estimates is applied. The essential condition for a.s. consistency at such x is given by

$$\sum_{i=1}^{\infty} \sum_{j=1}^{i} \phi(j)/(ik(i)) < \infty.$$

For points x which are atoms a direct approach is used which yields consistency under weak assumptions requiring only $k(n)/n \to 0$.

2. A GENERAL CONVERGENCE RESULT

The following general situation is considered. Let $Z, Z_1, Z_2, \ldots$ denote a sequence of random variables with common state space $\mathscr{S}$ and distributions $Q, Q_1, Q_2, \ldots$ on $\mathscr{S}$.

To treat possible nonstationarity we consider the total variation distance between Q_i and Q defined by

$$\delta(i) = \sup\{|Q_i(A) - Q(A)|: A \text{ measurable}\}.$$

It is well known that for any measurable bounded $f: \mathscr{S} \to (-\infty, +\infty)$ we have

$$\text{(S1)} \qquad |Ef(Z_i) - Ef(Z)| \leqslant 2\delta(i) \sup |f|.$$

Stationarity in the sense of $\delta(i) = 0$ can then be replaced by requiring $\delta(i)$ to converge to 0 suitably fast which holds for a large class of Markov processes irrespective of the initial distribution.

For $s > 1$ we shall also use the quantities

$$\delta_s(i) = \sup\{|\, \|f(Z_i)\|_s - \|f(Z)\|_s|: f \text{ measurable}, 0 \leqslant f \leqslant 1\}.$$

The mixing coefficients $\phi(i)$ and $\alpha(i)$ are defined by

$$\phi(i) = \sup\{|P(B \mid A) - P(B)|: B \in \sigma(Z_{k+i}), A \in \sigma(Z_k), k \geqslant 1\},$$

$$\alpha(i) = \sup\{|P(B \cap A) - P(B)\, P(A)|: B \in \sigma(Z_{k+i}), A \in \sigma(Z_k), k \geqslant 1\}.$$

Note that this relaxes the usual mixing conditions as it only makes use of $\sigma(Z_{k+i})$ instead of $\sigma(Z_j, j \geqslant k+i)$, resp. $\sigma(Z_k)$ instead of $\sigma(Z_j, j \geqslant k+i)$, resp. $\sigma(Z_k)$ instead of $\sigma(Z_j, j \leqslant k)$. For any measurable $f, g: \mathscr{S} \to (-\infty, +\infty)$ we have

$$\text{(S2)} \quad |Ef(Z_{i+n})\, g(Z_n) - Ef(Z_{i+n})\, Eg(Z_n)| \leqslant 2\phi(i)^{1/p}\, \|f(Z_{i+n})\|_q\, \|g(Z_n)\|_p,$$

$$\text{(S3)} \quad |Ef(Z_{i+n})\, g(Z_n) - Ef(Z_{i+n})\, Eg(Z_n)| \leqslant 8\alpha(i)^{1/r}\, \|f(Z_{i+n})\|_s\, \|g(Z_n)\|_t,$$

provided that the quantities on the right-hand side are defined and $1/p + 1/q = 1$, $1/r + 1/s + 1/t = 1$, $1 \leqslant p, q, r, s, t \leqslant \infty$.

Let $K_1, K_2, \ldots: \mathscr{S} \to (-\infty, +\infty)$ form a uniformly bounded sequence of measurable mappings.

Let $a(1), a(2), \ldots$ be a decreasing sequence of positive real numbers.

The following assumptions will play an important role for the convergence properties:

Assumption 1. $na(n) \to \infty$ as $n \to \infty$.

Assumption 2. $EK_n^{\pm}(Z)/a(n) = c^{\pm}(1+o(1))$ as $n \to \infty$ for some real numbers c^+, c^-, where $K_n^+ = \max\{K_n, 0\}$, $K_n^- = \max\{-K_n, 0\}$.

Assumption 3. $\sum_{i=1}^n \delta(i) = o(na(n))$ as $n \to \infty$.

Assumption 4. $\sum_{i=1}^n \phi(i) = o(na(n))$ as $n \to \infty$.

Assumption 5. There exists r, $1 \leqslant r < \infty$, such that with $s = r/(r-1)$

$$\sum_{i=1}^{n} \delta_s(i) = o(na(n)^{1/s}),$$

$$\sum_{i=1}^{n} \alpha(i)^{1/r} = o(na(n)^{1+1/r})$$

as $n \to \infty$.

Assumption 6. $\sum_{i=1}^\infty 1/(i^2 a(i)) < \infty$.

Assumption 7. $\lim_{n\to\infty} a([\gamma^n])/a([\gamma^{n+1}]) = A(\gamma) = 1 + o(1)$ as $\gamma \to 1$ with [ ] denoting integer part.

Assumption 8. $K_n^{\pm} \geqslant K_{n+1}^{\pm}$ for all n.

Assumption 9. $\sum_{i=1}^\infty \sum_{j=1}^i \phi(j)/(i^2 a(i)) < \infty$.

Assumption 10. There exists r, $1 \leqslant r < \infty$, such that

$$\sum_{i=1}^{\infty} \sum_{j=1}^{i} \alpha(j)^{1/r}/(i^2 a(i)^{1+1/r}) < \infty.$$

A detailed discussion of these assumptions will follow. Here we just give some general remarks.

Assumption 1 is indispensable in this field, assumption 6 follows from $na(n)/(\log n)^b \to \infty$ for some $b > 1$.

If we sequences $\delta(i)$ and $\phi(i)$ are summable then assumption 1 implies assumptions 3 and 4 and assumption 6 implies assumption 9 in the ϕ-mixing case. A similar conclusion holds of course in the α-mixing case.

Assumption 7 follows if for some monotone sequence $b(1) \leqslant b(2) \leqslant \ldots$ we have $na(n) \sim b(n)$.

If each K_n can be written in the form $K_n = \sum_{j=1}^J K_n^j$ for some J independent of n then it is sufficient to require assumptions 2 and 8 for each sequence K_n^j, $n = 1, 2, \ldots$ individually.

THEOREM 1. *Set*

$$V_n = 1/(na(n)) \sum_{i=1}^{n} (K_n(Z_i) - EK_n(Z_i)),$$

$$V_n' = 1/(na(n)) \sum_{i=1}^{n} (K_n(Z_i) - EK_n(Z)).$$

Then in the ϕ-mixing case

(i) *assumptions* 1–4 *imply*

$$V_n \to 0 \text{ in probability}, \qquad V_n' \to 0 \text{ in probability},$$

(ii) *assumptions* 1–4 *and* 6–9 *imply*

$$V_n \to 0 \ \text{a.s.}, \qquad V_n' \to 0 \ \text{a.s.}$$

and in the α-mixing case

(iii) *assumptions* 1–3, 5 *imply*

$$V_n \to 0 \text{ in probability}, \qquad V_n' \to 0 \text{ in probability},$$

(iv) *assumptions* 1–3, 5 *and* 6–8, 10 *imply*

$$V_n \to 0 \ \text{a.s.}, \qquad V_n' \to 0 \ \text{a.s.}$$

Proof. Let us look at the variances

$$v(n) = Var\, V_n.$$

We have

$$v(n) \leqslant g(n) + h(n),$$

where

$$g(n) = 1/(n^2 a(n)^2) \sum_{i=1}^{n} EK_n(Z_i)^2,$$

$$h(n) = 2/(n^2 a(n)^2) \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} Cov(K_n(Z_i), K_n(Z_j)).$$

Without loss of generality we may assume

$$K_n \geqslant 0 \qquad \text{for all } n.$$

In the following we let C denote a finite positive generic constant which may change at each appearance but is always independent of n.

(i) We consider the ϕ-mixing case. Then, using (S1)

$$g(n) \leqslant C/(n^2 a(n)^2) \sum_{i=1}^{n} \delta(i) + EK_n(Z)/(na(n)^2)$$

and with (S1) and (S2)

$$\begin{aligned} h(n) &\leqslant C/(n^2 a(n)^2) \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \phi(j-i)\, EK_n(Z_i) \\ &\leqslant C/(n^2 a(n)^2) \sum_{i=1}^{n-1} \phi(i) \sum_{j=1}^{n-i} \delta(j) + CEK_n(Z)/(n^2 a(n)^2) \sum_{i=1}^{n-1} \phi(i)(n-i) \\ &\leqslant (C/(na(n)) \sum_{i=1}^{n-1} \phi(i) + C(EK_n(Z)/a(n))/(na(n)) \sum_{i=1}^{n-1} \phi(i). \end{aligned}$$

Assumptions 1–4 immediately give $g(n) \to 0$ and $h(n) \to 0$ which proves part (i) of the results.

(ii) To obtain almost sure convergence, we follow a method used by Etemadi (1981) for a proof of the law of large numbers. Let $\gamma > 1$ and set

$$n_k = [\gamma^k].$$

Using the above inequalities for $g(n)$ and $h(n)$ assumptions 2, 5 and 6 show that

$$\sum_{i=1}^{\infty} (g(i) + h(i))/i < \infty,$$

which implies

$$\sum_{i=1}^{\infty} (g(n_i) + h(n_i)) = \sum_{i=1}^{\infty} v(n_i) < \infty.$$

This immediately yields

$$V_{n_k} \to 0 \text{ a.s.}$$

We now extend this convergence to the whole sequence V_n, $n = 1, 2, \ldots$ Consider n such that

$$n_k = [\gamma^k] \leqslant n \leqslant n_{k+1} = [\gamma^{k+1}].$$

We use $n_k = n'$, $n_{k+1} = n''$ for shorthand notation in the following estimates.

With assumption 8 we obtain

$$
\begin{aligned}
V_n &\leqslant 1/(n'a(n'')) \sum_{i=1}^{n''} K_{n'}(Z_i) - 1/(n''a(n')) \sum_{i=1}^{n'} EK_{n''}(Z_i) \\
&= W_n + d(n) + e(n),
\end{aligned}
$$

where

$$
\begin{aligned}
W_n &= 1/(n'a(n'')) \sum_{i=1}^{n''} (K_{n'}(Z_i) - EK_{n'}(Z_i)), \\
d(n) &= 1/(n'a(n'')) \sum_{i=1}^{n'} EK_{n'}(Z_i) - 1/(n''a(n')) \sum_{i=1}^{n'} EK_{n''}(Z_i), \\
e(n) &= 1/(n'a(n'')) \sum_{i=n'+1}^{n''} EK_{n'}(Z_i).
\end{aligned}
$$

Now

$$
W_n = (a(n')/a(n''))(n''/n')(1/(n''a(n'))) \sum_{i=1}^{n''} (K_{n'}(Z_i) - EK_{n'}(Z_i))
$$

so that

$$
W_n \to 0 \text{ a.s.}
$$

as in the first part of the proof with the additional use of assumption 7. Furthermore with (S1)

$$
\begin{aligned}
|d(n)| \leqslant {} & (a(n')/a(n''))(2/(n'a(n'))) \sum_{i=1}^{n'} \delta(i) \\
& + |(a(n')/a(n''))(EK_{n'}(Z)/a(n')) \\
& - (a(n'')/a(n'))(n'/n'')(EK_{n''}(Z)/a(n''))|,
\end{aligned}
$$

hence with assumptions 2 and 7

$$
\limsup_{n \to \infty} |d(n)| = B(\gamma),
$$

say, where

$$
B(\gamma) \to 0 \qquad \text{as} \quad \gamma \to 1.
$$

Similarly,

$$|e(n)| \leqslant (n''/n')(2/(n''a(n''))) \sum_{i=1}^{n''} \delta(i)$$
$$+ (a(n')/a(n''))((n''-n')/n'')(EK_{n'}(Z)/a(n')).$$

Thus

$$\limsup_{n\to\infty} |e(n)| = B'(\gamma),$$

say, where

$$B'(\gamma) \to 0 \qquad \text{as} \quad \gamma \to 1.$$

This implies for any $\gamma > 1$

$$\limsup_{n\to\infty} V_n \leqslant B(\gamma) + B'(\gamma) \text{ a.s.},$$

so that

$$\limsup_{n\to\infty} V_n \leqslant 0 \text{ a.s.}$$

In the same manner we use

$$V_n \geqslant 1/(n''a(n')) \sum_{i=1}^{n'} K_{n''}(Z_i) - 1/(n'a(n'')) \sum_{i=1}^{n''} EK_{n'}(Z_i)$$

to obtain

$$\liminf_{n\to\infty} V_n \geqslant 0 \text{ a.s.},$$

which proves the assertion for V_n.

Validity of the assertions for V_n' is immediate from this and the foregoing arguments.

(iii) The proof of (i), (ii) shows that we just have to look at $h(n)$ in the α-mixing case. Using r, s as in the assumptions we obtain with (S1) and (S3)

$$
\begin{aligned}
h(n) &\leqslant C/(n^2 a(n)^2) \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \alpha(j-i)^{1/r} \, \|K_n(Z_i)\|_s \\
&\leqslant C/(n^2 a(n)^2) \sum_{i=1}^{n-1} \alpha(i)^{1/r} \sum_{j=1}^{n-i} \delta_s(j) \\
&\quad + C(EK_n(Z))^{1/s}/(n^2 a(n)^2) \sum_{i=1}^{n-1} \alpha(i)^{1/r}(n-i) \\
&\leqslant C(EK_n(Z)/a(n))^{1/s}/(na(n)^{2-1/s}) \sum_{i=1}^{n-1} \alpha(i)^{1/r} \\
&\quad + (C/(na(n)^{2-1/s}) \sum_{i=1}^{n-1} \alpha(i)^{1/r}.
\end{aligned}
$$

Assumptions 1–3, 5 yield $h(n) \to 0$, so that convergence in probability follows.

(iv) Using assumption 10 now we obtain as in (ii) that

$$\sum_{i=1}^{\infty} (g(n_i) + h(n_i)) = \sum_{i=1}^{\infty} v(n_i) < \infty,$$

whence

$$V_{n_k} \to 0 \text{ a.s.}$$

The assertion of a.s. convergence then follows as in part (ii) of the proof.

The following consequences of Theorem 1 will only be stated for the case of a.s. convergence, as the conditions for convergence in probability do not seem to need further elaboration.

To arrive at a more concise although less general statement of the result we shall say that summability holds if

$$\sum_{i=1}^{\infty} (\delta(i) + \phi(i)) < \infty \qquad \text{for } \phi\text{-mixing},$$

resp.

$$\sum_{i=1}^{\infty} (\delta(i) + \delta_s(i) + \alpha(i)^{1/r}) < \infty \qquad \text{for } \alpha\text{-mixing}$$

in the setting of assumptions 3–5. So this notion of summability includes short-range dependence.

We then obtain the following simpler version.

Corollary 1. *Assume summability and let assumptions* 1, 2 *and* 7, 8 *be valid. Then*

(i) *in the* ϕ*-mixing case*

$$\sum_{i=1}^{\infty} 1/(i^2 a(i)) < \infty \qquad \textit{implies} \ \ V_n \to 0 \ \ a.s., \ \ V'_n \to 0 \ \ a.s.;$$

(ii) *in the* α*-mixing case*

$$\sum_{i=1}^{\infty} 1/(i^2 a(i)^{1+1/r}) < \infty \qquad \textit{implies} \quad V_n \to 0 \ \ a.s., \ \ V'_n \to 0 \ \ a.s.,$$

where r, $1 \leqslant r < \infty$, *arises from the summability assumption.*

This follows immediately from Theorem 1.

Remark 1. (i) For ϕ-mixing, results in the spirit of Theorem 1 are given by Collomb (1984), Lemma 3, see also Györfi, Härdle, Sarda and Vieu (1989), Section III.3.3, and Roussas (1988), Theorems 2.1, 2.2. Both authors use exponential probabilistic inequalities and arrive at a.s. convergence via complete convergence and the Borel-Cantelli Lemma.

Collomb (1984) uses a condition on the interplay between $\phi(n)$ and $a(n)$ which in the case $\phi(n) \sim ab^n$, $0 < b < 1$, holds if

$$na(n)/(\log n)^2 \to \infty,$$

and in the case $\phi(n) \sim an^{-b}$, $b > 1$, holds if

$$na(n)/(n^{1/(1+b)} \log n) \to \infty.$$

Our result requires in these situations only

$$na(n)/(\log n)^c \to \infty \qquad \text{for some} \quad c > 1,$$

summability being obviously fulfilled.

Roussas (1988) obtains his results under the assumption of summability. His conditions can be compaired to ours in the case $a(n) = n^{-c}$. In this case he requires $c < 1/2$ whereas our result holds for $c < 1$.

(ii) For α-mixing, we refer to Györfi, Härdle, Sarda and Vieu (1989), Section III.3.5 and again Roussas (1988), Theorems 2.1, 2.2. In the first reference an involved condition on the interplay between $\alpha(n)$ and $a(n)$ is used which in the case $\alpha(n) \sim ab^n$, $0 < b < 1$, holds if

$$na(n)/(n^{1/2} \log n) \to \infty.$$

For this case of geometrically decreasing mixing coefficients we may use in the summability condition $1/r$ arbitrarily close to 0, hence $1+1/r$ arbitrarily close to 1, so that our result holds if

$$na(n)/n^c \to \infty \qquad \text{for some} \quad c>0.$$

Comparing the results of Roussas (1988) to ours for $a(n)=n^{-c}$ he requires $c<1/2$, whereas our result holds for $c<1$. In the case $\alpha(n) \sim an^{-b}$, $b>1$, he uses

$$a(n)=n^{-c} \qquad \text{with} \quad c<(b-1)/b$$

compared to

$$a(n)=n^{-c} \qquad \text{with} \quad c<b/(b+1)$$

in our result. For polynomially decreasing α-mixing coefficients the conditions of Györfi, Härdle, Sarda and Vieu (1989) are not fulfilled for any power $a(n)=n^{-c}$.

(iii) These comparisons show that for a.s. convergence the method of Etemadi (1981) applied to this context leads to substantially more general conditions and avoids the technical difficulties of exponential probabilistic inequalities under mixing. It should be noted that ϕ-mixing and α-mixing are treated in an essentially unified manner. As mixing only enters via inequalities for covariances other concepts of mixing and dependence can be treated along the same lines. From Berbee (1987) it may be seen that our assumptions in general do not provide complete convergence.

(iv) In the aforementioned references the question of uniform a.s. convergence is then treated for kernel estimates. Here the method of exponential probabilistic inequalities shows its full power and yields less restrictive conditions than the method proposed in this paper, which seems to be particularly suited to obtain pointwise a.s. convergence.

In order to avoid overburdening technicalities the foregoing results were given under the condition of uniform boundedness. We shall next formulate a result for the unbounded case assuming stationarity in the sense of $Q=Q_n$ for all n.

COROLLARY 2. *Assume $Q=Q_n$ for all n and validity of assumptions* 1, 6–8. *Let $f\colon \mathscr{S} \to (-\infty, +\infty)$ and set*

$$W_n = 1/(na(n)) \sum_{i=1}^{n} (f(Z_i)\, K_n(Z_i) - Ef(Z)\, K_n(Z)).$$

(i) *In the ϕ-mixing case let $p \geqslant 2$ and $q = p/(p-1)$ and assume that $E\,|f(Z)|^p < \infty$, $Eg(Z)\,K_n(Z)^{\pm}/a(n)$ converges to a finite limit as $n \to \infty$ for $g = (f^{\pm})^t$, $t = 1, 2, q, p$. Then*

$$\sum_{i=1}^{\infty} \sum_{j=1}^{i} \phi(j)^{1/q}/(i^2 a(i)) < \infty \qquad \textit{implies} \quad W_n \to 0 \ \textit{a.s.}$$

(ii) *In the α-mixing case let $s > 2$ and $r = s/(s-2)$ and assume that $E\,|f(Z)|^s < \infty$, $Eg(Z)\,K_n(Z)^{\pm}/a(n)$ converges to a finite limit as $n \to \infty$ for $g = (f^{\pm})^t$, $t = 1, 2, s$. Then*

$$\sum_{i=1}^{\infty} \sum_{j=1}^{i} \alpha(j)^{1/r}/(i^2 a(i)^{1+1/r}) < \infty \qquad \textit{implies} \quad W_n \to 0 \ \textit{a.s.}$$

Proof. (i) In the ϕ-mixing case we use $\delta(i) = 0$ and uniform boundedness of K_n to obtain with (S2)

$$\begin{aligned} g(n) &\leqslant CEf(Z)^2\,K_n(Z)/na(n)^2, \\ h(n) &\leqslant C/(n^2 a(n)^2) \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \phi(j-i)^{1/q}\,\|f(Z)\,K_n(Z)\|_p\,\|f(Z)\,K_n(Z)\|_q \\ &\leqslant C(Ef(Z)^p K_n(Z)/a(n))^{1/p}(Ef(Z)^q K_n(Z)/a(n))^{1/q} \\ &\quad \times (1/na(n)) \sum_{i=1}^{n-1} \phi(i)^{1/q}. \end{aligned}$$

(ii) In the α-mixing case it is enough to point out that with (S3)

$$h(n) \leqslant C(Ef(Z)^s K_n(Z)/a(n))^{2/s}(1/na(n)^{2-2/s}) \sum_{i=1}^{n-1} \alpha(i)^{1/r}.$$

Remark 2. The unbounded case is also covered in III.3.3 and III.3.5 of Györfi, Härdle, Sarda and Vieu (1989) under the assumption $E\,|f(Z)|^p < \infty$ for some $p > 2$. A cutoff sequence $M_n = n^t$, $t > 4/(p+2)$ is used.

(i) The ϕ-mixing case is treated in Section III.3.3 of the above reference and the condition for the result becomes in the case $\phi(n) \sim ab^n$, $0 < b < 1$,

$$na(n)/(n^t(\log n)^2) \to \infty,$$

and in the case $\phi(n) \sim an^{-b}$, $b > 1$,

$$na(n)/(n^{t+1/(1+b)} \log n) \to \infty, \qquad p > 2 + 4/b.$$

Our result requires in these situations

$$na(n)/(\log n)^c \to \infty \qquad \text{for some} \quad c > 1,$$

and

$$E\,|f(Z)|^p < \infty \qquad \text{for} \quad p = 2, \qquad \text{resp. } p > b/(b-1).$$

(ii) For the α-mixing case as treated in Section III.3.5 the requirement when $\alpha(n) \sim ab^n$, $0 < b < 1$, becomes

$$na(n)/(n^{t+1/2}(\log n)^2) \to \infty, \qquad p > 6.$$

It cannot be fulfilled for $\alpha(n) \sim an^{-b}$, $b > 1$.

In the geometrically decreasing case Corollary 2 requires

$$na(n)/n^c \to \infty \qquad \text{for some} \quad c > 0$$

and

$$E\,|f(Z)|^p < \infty \qquad \text{for some} \quad p > 2.$$

In the case $\alpha(n) \sim an^{-b}$, $b > 1$,

$$na(n)/n^c \to \infty \qquad \text{for some} \quad c > 1/(b+1)$$

and

$$E\,|f(Z)|^p < \infty \qquad \text{for some} \quad p > 2b/(b-1).$$

3. KERNEL ESTIMATION

Density Estimation

We start with the problem of density estimation and let $\mathscr{S}$ be euclidean space $(-\infty, +\infty)^d$. Let furthermore

$$K\colon (-\infty, +\infty)^d \to [0, \infty), \qquad \int K(x)\, \lambda(dx) = 1,$$

such that

$$K(xh) \geqslant K(x) \qquad \text{for all real } x,\ 0 \leqslant h \leqslant 1.$$

Here λ denotes d-dimensional Lebesgue-measure. For a decreasing sequence $h(1)$, $h(2)$, ... of positive real numbers we set

$$K_n(x, z) = K((x-z)/h(n)),\ x, z \in (-\infty, +\infty)^d;$$

furthermore $a(n) = h(n)^d$ and

$$V_n'(x) = (1/nh(n)^d) \sum_{i=1}^{n} (K_n(x, Z_i) - EK_n(x, Z))$$

so that the kernel density estimator becomes

$$f_n(x) = V_n'(x) + EK_n(x, Z)/h(n)^d.$$

Let us look at the validity of our assumptions in this example.

Assumption 8 is fulfilled since K is assumed decreasing in h. If the sequence $nh(n)^d$ is increasing and converges to ∞, then assumptions 1 and 7 follow.

Here assumption 2 is a thoroughly investigated topic if the distribution Q of Z has a Lebesgue density f. Conditions for the validity of

$$\lim_{h \to 0} EK((x-Z)/h))/h^d = f(x) \qquad \text{for a.a. } x,$$

are well-known, e.g. Wheeden and Zygmund (1977), Theorem (9.13). Here we shall simply call K regular if the above convergence holds.

Hence we arrive at the following corollary which we shall only state for the summable case.

COROLLARY 3. *Assume summability and let the distribution Q of Z have a Lebesgue density f. Assume that $nh(n)^d$ is increasing and tends to ∞ and that K is regular. Then*

(i) *in the ϕ-mixing case*

$$\sum_{i=1}^{\infty} 1/(i^2 h(i)^d) < \infty \qquad \textit{implies } f_n(x) \to f(x) \textit{ a.s. for a.a. } x;$$

(ii) *in the α-mixing case*

$$\sum_{i=1}^{\infty} 1/(i^2 h(i)^{d+d/r}) < \infty \qquad \textit{implies } f_n(x) \to f(x) \textit{ a.s. for a.a } x,$$

where r, $1 \leqslant r < \infty$, arises from the summability assumption.

Remark 3. In density estimation kernels taking positive and negative values are used. Such kernels will in general not satisfy the assumption

$$K(xh) \geqslant K(x) \qquad \text{for all real } x,\ 0 \leqslant h \leqslant 1.$$

We shall now show how to treat this more general case.

We start with the univariate case and let

$$K: (-\infty, +\infty) \to (-\infty, \infty).$$

Assume that K is absolutely continuous which implies the existence of K' almost everywhere with respect to Lebesgue measure and that

$$\int |K'(t)|\, dt < \infty, \qquad \lim_{x \to \pm\infty} K(x) = 0.$$

Consider firstly K on $[0, \infty)$.

Following Wheeden and Zygmund (1977), (7.31), we obtain

$$K(x) = K_1(x) - K_2(x),$$

where $K_1(x)$, $K_2(x)$ are decreasing in $x \in [0, \infty)$ and $\geqslant 0$. If K has bounded support then this is also true for K_1, K_2.

Similarly we may represent K on $(-\infty, 0)$ as difference $K(x) = K_1(x) - K_2(x)$ of two increasing functions on $(-\infty, 0)$. Thus we arrive at $K(x) = K_1(x) - K_2(x)$ on $(-\infty, +\infty)$ with each $K_i \geqslant 0$ and satisfying

$$K_i(xh) \geqslant K_i(x) \qquad \text{for all real } x,\ 0 \leqslant h \leqslant 1.$$

So we may apply the methods of this paper separately to K_1, K_2 to obtain the required results for general K.

A similar argument applies to multivariate kernels of the form

$$K(x_1, ..., x_n) = \prod_{j=1}^{n} K_j(x_j),$$

representing each of the univariate kernels as above.

Regression Estimation

We next turn to regression estimation and let $\mathcal{S} = (-\infty, +\infty)^d \times \mathcal{Y}$, furthermore $Z = (X, Y)$, $Z_1 = (X_1, Y_1)$, ... Let $f: \mathcal{Y} \to (-\infty, +\infty)$ and set

$$r(x) = E(f(Y) \mid X = x)$$

assuming integrability of $f(Y)$. As above let

$$K: (-\infty, +\infty)^d \to [0, \infty)$$

such that

$$K(xh) \geqslant K(x) \qquad \text{for all} \quad 0 \leqslant h \leqslant 1.$$

For a decreasing sequence $h(1), h(2), \dots$ of positive real numbers we set

$$K_n(x, z) = K((x - z)/h(n)), \ x, z \in (-\infty, +\infty)^d;$$

furthermore $a(n, x) = EK_n(x, X)$. Let

$$V_n^j(x) = (1/na(n, x)) \sum_{i=1}^{n} (f(Y_i)^j K_n(x, X_i) - Ef(Y)^j K_n(x, X)) \quad \text{for} \quad j = 0, 1.$$

Then the kernel regression estimator becomes

$$r_n(x) = (V_n^1(x) + Ef(Y) K_n(x, X)/EK_n(x, X))/(V_n^0(x) + 1).$$

Let us look at the validity of our assumptions for this situation.

Assumption 8 is fulfilled as above.

From Krzyzak and Pawlak (1984) we have

$$\lim_{h \to 0} Ef(Y) K((x - X)/h)/EK((x - X)/h) = E(f(Y) \mid X = x) \qquad \text{for a.a. } x$$

provided that K has bounded support and $K(x) \geqslant c$ for $\|x\| \leqslant t_0$ for some $t_0, c > 0$. We shall simply call K regression regular if the above convergence holds. Thus assumption 2, resp. the corresponding assumptions in Corollary 2 hold in this situation.

For a treatment of the other assumptions we employ a result from Devroye (1981). Letting Q' denote the distribution of X and $B(x, h)$ the open or closed ball with center x and radius h this states

$$Q'(B(x, h))/h^d \to g(x) \qquad \text{for } Q'\text{-a.a. } x \text{ as } h \to 0,$$

where g is a measurable mapping satisfying

$$0 < g(x) \leqslant \infty \qquad \text{for all } x.$$

This implies that, $a(n, x) = EK_n(x, X)$ fulfills any of assumptions j, j = 1, 3–6, 9, 10, if $h(n)^d$ fulfills the corresponding assumption.

The situation for assumption 7 seems to be more delicate. As pointed out this assumption holds if

$$na(n, x) \sim b(n, x) \qquad \text{for some increasing sequence } b(n, x), \ n = 1, 2, \dots .$$

If Q' has a Lebesgue density f and $nh(n)^d$ is increasing then we may use $b(n, x) = \int K(x)\, \lambda(dx)\, f(x)\, nh(n)^d$ to obtain assumption 7 for any x such that $f(x) > 0$ hence for Q'-a.a. x.

If X has a discrete distribution then we may simply choose $b(n, x) = K(x)\, nP(X = x)$ to obtain assumption 7 for Q'-a.a. x. Note that in our arguments we do not require $a(n) \to 0$ which of course fails in the discrete case.

We do not have any general result available to verify assumption 7. As a shorthand expression let us call $x \in (-\infty, +\infty)^d$ regular if assumption 7 holds for $a(n, x)$, hence Q'-a.a. points are regular in the case of Lebesgue absolutely continuous and of discrete Q'. Of course there is a wealth of other distributions such that a.a. points are regular.

We then arrive at the following result on a.s. consistency of kernel regression estimators, where we only give a formulation for the case of bounded f. This includes the result of Theorem 1 of Collomb (1984) for a.s. convergence under more general conditions, see the discussion in Remark 1.

COROLLARY 4. *Assume that $nh(n)^d$ tends to ∞ and K is regression regular. Then*

(i) *in the ϕ-mixing case assumptions* 3, 4, 6 *and* 9 *for* $a(n) = h(n)^d$ *imply*

$$r_n(x) \to r(x) \text{ a.s.} \qquad \text{for } Q'\text{-a.a. regular } x,$$

(ii) *in the α-mixing case assumptions* 3, 5, 6 *and* 10 *for* $a(n) = h(n)^d$ *imply*

$$r_n(x) \to r(x) \text{ a.s.} \qquad \text{for } Q'\text{-a.a. regular } x.$$

Thus under summability the above convergence at regular points is implied by

$$\sum_{i=1}^{\infty} 1/(i^2 h(i)^{d'}) < \infty$$

with $d' = d$ for ϕ-mixing, $d' = d + d/r$ for α-mixing. The case of unbounded f can be treated as in Corollary 2 under stationarity. The formulation is straightforward and omitted.

4. NEAREST NEIGHBOUR ESTIMATION

In the problem of estimating a d-dimensional Lebesgue density Moore and Yackel (1977) have shown that consistency of kernel density estimtors leads to the corresponding results on consistency for nearest neighbour estimators. We shall use their approach in the context of nonparametric regression for the case that $P(X=x)=0$. The nearest neighbour regression estimator will be investigated by direct means if $P(X=x)>0$.

We thus consider regression estimation and as in section 3 let $\mathscr{S}=(-\infty, +\infty)^d \times \mathscr{Y}$, $Z=(X, Y)$, $Z_1=(X_1, Y_1)$, ...

Let $f\colon \mathscr{Y} \to (-\infty, +\infty)$ and set

$$r(x)=E(f(Y) \mid X=x$$

assuming integrability of $f(Y)$.

Let $x \in (-\infty, +\infty)^d$. Rank (X_i, Y_i) according to increasing values of $\|X_i - x\|$ where ties are broken by comparing indices. We thus obtain a random permutation

$$R_i = R_i^n(x, X_1, ..., X_n), \qquad i=1, ..., n$$

of the indices $i=1, ..., n$ such that

$$\|X_{R_1} - x\| \leqslant \|X_{R_2} - x\| \leqslant \cdots \leqslant \|X_{R_n} - x\|$$

and

$$\|X_{R_i} - x\| = \|X_{R_{i+1}} - x\| \qquad \text{implies } R_i < R_{i+1}.$$

We remark that in the case $P(X=x)=0$ the following arguments are valid for any tie breaking rule. Let $k(n)$, $n=1, 2, ...$ be an increasing sequence of integers >0. Let

$$D_n(x) = \|X_{R_{k(n)}} - x\|$$

denote the distance from x to the $k(n)$-nearest neighbour $X_{R_{k(n)}}$. Let $I(x, n) \subseteq \{1, ..., n\}$ be a random set of indices. Then

$$m_n(x) = 1/k(n) \sum_{i \in I(x,n)} f(Y_i)$$

with

$$\{i\colon \|X_i - x\| < D_n(x)\} \subseteq I(x, n) \subseteq \{i\colon \|X_i - x\| \leqslant D_n(x)\} \qquad \text{if } P(X=x)=0,$$

$$I(x, n) = \{R_1(x), ..., R_{k(n)}(x)\} \qquad \text{if} \quad P(X=x)>0,$$

is called a $k(n)$-nearest neighbour regression estimator.

In the case $P(X=x)=0$ this is clearly independent of the tie breaking rule.

In the following, F_x denotes the distribution function of $\|X-x\|$, so that

$$F_x(h)=Q'(S(x,h)), \qquad F_x(h-)=Q'(S^o(x,h))$$

for the closed ball $S(x,h)$ and the open ball $S^o(x,h)$ with center x and radius h.

In the next lemma, the indicator function of a set A will be denoted by 1(A).

LEMMA 1. *Let $x\in(-\infty,+\infty)^d$ such that $F_x(0)=0$, $\lim_{h\to 0} F_x(h-)/F_x(h)=1$. Let $h(n,x)=F_x^{-1}(k(n)/n)$ and*

$$U_n(x)=1/k(n)\sum_{i=1}^{n} 1(X_i\in B(x,h(n,x)))$$

with B denoting the open or closed ball. Assume that $k(n)$ is increasing with $\lim_{n\to\infty} k(n)=\infty$, and $k(n)/n$ is decreasing with $\lim_{n\to\infty} k(n)/n=0$. Then in the ϕ-mixing case assumptions 3, 4, 6 *and* 9 *imply*

$$U_n(x)\to 1 \ \ a.s.,$$

and in the α-mixing case assumptions 3, 5, 6 *and* 10 *imply*

$$U_n(x)\to 1 \ \ a.s.$$

Proof. We apply Theorem 1 to $a(n)=k(n)/n$ and $K_n(Z_i)=1\ (X_i\in B(x,h(n,x)))$. Note that assumption 7 is fulfilled since $a(n)=k(n)/n$ is decreasing and $na(n)=k(n)$ is increasing. It remains to look at assumption 2. Obviously

$$\begin{aligned}|EK_n(Z)/a(n)-1| &= |(F_x(F_x^{-1}(a(n)))-a(n))/a(n)|\\ &\leqslant (F_x(h(n,x))-F_x(h(n,x)-))/F_x(h(n,x)-)),\end{aligned}$$

where by assumption

$$\lim_{n\to\infty}(F_x(h(n,x))-F_x(h(n,x)-))/F_x(h(n,x)-))=0$$

and the assertions then follow from Theorem 1.

LEMMA 2. *Let $x\in(-\infty,+\infty)^d$. For $c>0$ set $a(n,c)=ck(n)/n$, $h(n,x,c)=F_x^{-1}(a(n,c))$ and*

$$U_n(x,c)=1/na(n,c)\sum_{i=1}^{n} 1(X_i\in B(x,h(n,x,c)))$$

with B denoting the open or closed ball. Then $U_n(x, c) \to 1$ a.s. for open and closed B and all $c > 0$ implies

$$F_x(D_n(x))/(k(n)/n) \to 1 \ \text{a.s.}$$

Proof. We omit the dependence on x in our notations for this proof. Let $\varepsilon > 0$. Then

$$F(D_n) > (1+\varepsilon)\, k(n)/n = a(n, 1+\varepsilon)$$

implies

$$D_n \geqslant F^{-1}(a(n, 1+\varepsilon)) = h(n, 1+\varepsilon);$$

hence

$$k(n) \geqslant \sum_{i=1}^{n} 1(X_i \in S^o(x, h(n, 1+\varepsilon)))$$

and

$$k(n)/(na(n, 1+\varepsilon)) = 1/(1+\varepsilon) \geqslant 1/(na(n, 1+\varepsilon)) \sum_{i=1}^{n} 1(X_i \in S^o(x, h(n, 1+\varepsilon))).$$

Similarly

$$F(D_n) < (1-\varepsilon)\, k(n)/n = a(n, 1-\varepsilon)$$

yields

$$D_n < F^{-1}(a(n, 1-\varepsilon)) = h(n, 1-\varepsilon);$$

thus

$$k(n)/(na(n, 1-\varepsilon)) = 1/(1-\varepsilon) \leqslant 1/(na(n, 1-\varepsilon)) \sum_{i=1}^{n} 1(X_i \in S(x, h(n, 1-\varepsilon))).$$

This immediately implies the assertion.

We have now all the tools to obtain consistency for points x with $P(X = x) = 0$ under an additional requirement as specified in Lemma 1.

THEOREM 2. *Let f be bounded. Assume that $k(n)$ is increasing with $\lim_{n \to \infty} k(n) = \infty$, and $k(n)/n$ is decreasing with $\lim_{n \to \infty} k(n)/n = 0$. For*

Q'-a.a. $x \in (-\infty, +\infty)^d$ with $F_x(0)=0$, $\lim_{h\to 0} F_x(h-)/F_x(h)=1$ the following holds: In the ϕ-mixing case assumptions 3, 4, 6 *and* 9 *imply*

$$m_n(x) \to r(x) \text{ a.s.},$$

and in the α-mixing case assumptions 3, 5, 6 *and* 10 *imply*

$$m_n(x) \to r(x) \text{ a.s.}$$

Proof. We assume without loss of generality $f \geqslant 0$ and omit the dependence on x in our notations.

Let $\delta > 0$ and with $a(n, c)$ and $h(n, c)$ as in Lemma 2 set

$$A_n = \{a(n, 1-\delta) < F(D_n) < a(n, 1+\delta)\}.$$

Then on A_n we have $h(n, 1-\delta) \leqslant D_n < h(n, 1+\delta)$ and

$$\sum_{i=1}^{n} f(Y_i)\, 1(X_i \in S^o(x, h(n, 1-\delta)))$$
$$\leqslant k(n)\, m_n \leqslant \sum_{i=1}^{n} f(Y_i)\, 1(X_i \in S(x, h(n, 1+\delta)));$$

furthermore,

$$\sum_{i=1}^{n} 1(X_i \in S^o(x, h(n, 1-\delta))) \leqslant k(n) \leqslant \sum_{i=1}^{n} 1(X_i \in S(x, h(n, 1+\delta))).$$

Letting

$$r_n^0 = \sum_{i=1}^{n} f(Y_i)\, 1(X_i \in S^o(x, h(n, 1-\delta))) \Big/ \sum_{i=1}^{n} 1(X_i \in S^o(x, h(n, 1-\delta))),$$
$$r_n^1 = \sum_{i=1}^{n} f(Y_i)\, 1(X_i \in S(x, h(n, 1+\delta))) \Big/ \sum_{i=1}^{n} 1(X_i \in S(x, h(n, 1+\delta))),$$

and

$$U_n^0 = \sum_{i=1}^{n} 1(X_i \in S^o(x, h(n, 1-\delta)))/(na(n, 1-\delta)),$$
$$U_n^1 = \sum_{i=1}^{n} 1(X_i \in S(x, h(n, 1+\delta)))/(na(n, 1+\delta))$$

we obtain on A_n,

$$r_n^0(U_n^0/U_n^1)(1-\delta)/(1+\delta) \leqslant m_n \leqslant r_n^1(U_n^1/U_n^0)(1+\delta)/(1-\delta).$$

The assertions of the theorem will obviously follow if

$$F(D_n)/(k(n)/n) \to 1, \qquad U_n^j \to 1, r_n^j \to r(x),\ j=0, 1,$$

holds a.s. under the assumptions as given above.

Now for $F(D_n)/(k(n)/n)$ and U_n^j, $j=0, 1$, this is immediate from Lemmas 1 and 2. For r_n^j, $j=0, 1$, this follows from Theorem 1 as in Corollary 4 noting that

$$nE1(X_i \in S^o(x, h(n, 1-\delta)) \sim (1-\delta)\, k(n),$$

$$nE1(X_i \in S(x, h(n, 1+\delta)) \sim (1+\delta)\, k(n).$$

Under summability we arrive at the sufficient condition

$$\sum_{i=1}^{\infty} 1/(i^{1-c}k(i)^{1+c}) < \infty$$

with $c=0$ for ϕ-mixing and $c=1/r$ for α-mixing.

In the following we shall see that consistency for nearest neighbour regression estimators holds under nonrestrictive conditions at a point x with $P(X=x)>0$. Thus in order not to have consistency for Q'-a.a. x we need an uncountable set of points x such that $P(X=x)=0$ and $\lim_{h\to 0} F_x(h-)/F_x(h)<1$. In particular there has to exist to any of these uncountably many x a sequence $h(n, x)$, $n=1, 2, \ldots$ converging to 0 such that $P(X=h(n, x))>0$. It does not seem to be obvious whether such a distribution may exist.

We now turn to the case that $P(X=x)>0$. The following lemma which may be of some independent interest will imply the desired result. As in Section 2 we firstly consider a general sequence $Z_1, Z_2, \ldots$ of random variables with common state space $\mathscr{S}$ equipped with a σ-algebra $\mathscr{A}$.

LEMMA 3. *Let F be a set of measurable real-valued functions on $\mathscr{S}$ such that $\{1(A)\colon A\in\mathscr{A}\}\subseteq\mathscr{F}$ and $f1(A)\in F$ for all $f\in F$ and $A\in\mathscr{A}$. Assume that W is a probability measure on $\mathscr{S}$ such that for all $f\in F$,*

$$1/n \sum_{i=1}^{n} f(Z_i) \to \int f\,dW \ \textit{a.s.}$$

For $B\in\mathscr{A}$ with $W(B)>0$ define

$$T(n)=\inf\{k>T(n-1)\colon Z_k\in B\} \qquad \textit{with} \quad T(0)=0.$$

Then $T(n) < \infty$ a.s. for all n and

$$1/n \sum_{i=1}^{n} f(Z_{T(i)}) \to \int_B fdW/W(B) \text{ a.s.}$$

Proof. Let $M = \{Z_n \in B \text{ infinitely often}\}$ hence $T(n) < \infty$ for all n on M. Furthermore

$$1/n \sum_{i=1}^{n} 1(Z_i \in B) \to 0 \qquad \text{on } M^c,$$

which implies $P(M^c) = 0$ since

$$1/n \sum_{i=1}^{n} 1(Z_i \in B) \to W(B) > 0 \text{ a.s.}$$

Let $f \in F$. Then

$$1/n \sum_{i=1}^{n} f(Z_i)\, 1(Z_i \in B) \to \int_B fdW,$$

$$1/n \sum_{i=1}^{n} 1(Z_i \in B) \to W(B)$$

imply

$$1/T(n) \sum_{i=1}^{T(n)} f(Z_i)\, 1(Z_i \in B) \to \int_B fdW,$$

$$1/T(n) \sum_{i=1}^{T(n)} 1(Z_i \in B) \to W(B);$$

hence

$$\sum_{i=1}^{T(n)} f(Z_i)\, 1(Z_i \in B) \Big/ \sum_{i=1}^{T(n)} 1(Z_i \in B) \to \int_B fdW/W(B).$$

The assertion then follows from the identity

$$\sum_{i=1}^{T(n)} f(Z_i)\, 1(Z_i \in B) \Big/ \sum_{i=1}^{T(n)} 1(Z_i \in B) = \sum_{i=1}^{n} f(Z_{T(i)})/n.$$

In the Markovian setting, $Z_{T(n)}$ is called the process on B and this lemma identifies in a simple way the stationary distribution of the process on B as the stationary distribution of the original process conditioned on B, a result usually proved by different means.

We now go back to the situation

$$\mathscr{S} = (-\infty, +\infty)^d \times \mathscr{Y}, \qquad Z = (X, Y), \qquad Z_1 = (X_1, Y_1), \ldots.$$

For f: $\mathscr{S}(-\infty, +\infty)$ set

$$m_n(f, x) = 1/k(n) \sum_{i \in I(x,n)} f(Z_i),$$

where $I(x, n) = \{R_1(x), \ldots, R_{k(n)}(x)\}$.

THEOREM 3. *Let F be a set of measurable real-valued functions on $\mathscr{S}$ such that $\{1(A)\colon A \in \mathscr{A}\} \subseteq \mathscr{F}$ and $f1(A) \in F$ for all $f \in F$ and $A \in \mathscr{A}$. Assume that W is a probability measure on $\mathscr{S}$ such that for all $f \in F$*

$$1/n \sum_{i=1}^{n} f(Z_i) \to \int f dW \ \textit{a.s.}$$

Let $x \in (-\infty, +\infty)^d$ with $W(\{x\} \times \mathscr{Y}) > 0$. Then for every $f \in F$

$$m_n(f, x) \to \int_{\{x\} \times \mathscr{Y}} f dW / W(\{x\} \times \mathscr{Y}) \ \textit{a.s.}$$

provided that $k(n) \to \infty$ and $k(n)/n \to 0$.

Proof. Let $D_n = D_n(x) = \|X_{R_{k(n)}} - x\|$ denote the distance from x to the $k(n)$-nearest neighbour $X_{R_{k(n)}}$. Setting $M = \{D_n = 0 \text{ for almost all } n\}$ we obtain

$$\begin{aligned} M^c &= \{D_n > 0 \text{ for infinitely many } n\} \\ &= \left\{1/n \sum_{i=1}^{n} 1(X_i = x) < k(n)/n \text{ for infinitely many } n\right\} \\ &\subseteq \left\{1/n \sum_{i=1}^{n} 1(Z_i \in \{x\} \times \mathscr{Y}) < W(\{x\} \times \mathscr{Y})/2 \text{ for infinitely many } n\right\}. \end{aligned}$$

But the probability of the last event is equal to 0 since

$$1/n \sum_{i=1}^{n} 1(Z_i \in \{x\} \times \mathscr{Y}) \to W(\{x\} \times \mathscr{Y}) > 0$$

so that we obtain

$$P(D_n = 0 \text{ for almost all } n) = 1.$$

Set

$$T(n) = \inf\{k > T(n-1)\colon Z_k \in \{x\} \times \mathscr{Y}\}$$

with $T(0)=0$. Then $D_n=0$ implies

$$m_n(f,x)=1/k(n)\sum_{i=1}^{k(n)} f(Z_{T(i)}),$$

where we used the tie breaking rule as described in the beginning of the section. The assertion thus follows from Lemma 3.

For W being the distribution of $Z=(X, Y)$ we may simply write

$$\int_{\{x\}\times\mathcal{Y}} f dW/W(\{x\}\times\mathcal{Y})=E(f(X, Y)\mid X=x)$$

so that consistency of nearest neighbour regression estimators at x with $P(X=x)>0$ is implied by the validity of a law of large numbers. Results of this type under mixing conditions are well-known, see e.g. Berbee (1987) where for bounded f such a result is shown under the assumption $\sum_{i=1}^{\infty}\alpha(i)/i<\infty$.

ACKNOWLEDGMENT

The author thanks the Centre for Mathematics and Its Applications, School of Mathematical Sciences, Australian National University, where he did this work as a Visiting Fellow. He was supported by a grant of the Volkswagenstiftung.

REFERENCES

Berbee, H. (1987). Convergence rates in the strong law for bounded mixing sequences. *Probab. Theory Relat. Fields* **74** 255–270.

Collomb, G. (1984). Propriétés de convergence presque complète du prédicteur a noyau. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **66** 441–460.

Devroye, L. (1981). On the almost everywhere convergence of nonparametric regression function estimates. *Ann. Statist.* **9** 1310–1319.

Etemadi, N. (1981). An elementary proof of the law of large numbers. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **55** 119–122.

Györfi, L., Härdle, W., Sarda, P., and Vieu, P. (1989). *Nonparametric Curve Estimation from Time Series*, Lecture Notes in Statististics, Vol. 60. Springer-Verlag, Berlin.

Krzyzak, A., and Pawlak, M. (1984). Distribution-free consistency of a nonparametric kernel regression estimate and classification. *IEEE Trans. Inform. Theory* **30** 78–81.

Moore, D. S., and Yackel, Y. W. (1977). Consistency properties of nearest neighbor density function estimators. *Ann. Statist.* **5** 143–154.

Roussas, G. G. (1988). Nonparametric estimation in mixing sequences of random variables. *J. Statist. Plan. Inference* **18** 135–149.

Wheeden, R. L., and Zygmund, A. (1977). *Measure and Integral.* Dekker, New York.